REGIONAL STATISTICS AND THE INDIAN NATIONAL SAMPLE SURVEY*

By D. B. LAHIRI

Indian Statistical Institute

SUMMARY. This paper gives more or less chronologically an account of the thoughts, motivations and actions of the Indian National Sample Survey (NSS) about the production of statistics for regions of different size. It explains how the main features of the NSS, namely, multi-agency, multipurpose, multi-subject and continuing, have been exploited to secure data for greater and greater number of regions, and at the same time, to obtain more precise and useful information. It also describes how the NSS has coordinated technically, and synchronized operationally, activities of different regional authorities, with a view to obtaining results, comparable but independent. This, with similar but intra-agency inter-penetrating network of samples has provided a means to test the validity of the results, and also to assess the uncertainty in the estimates. It enumerates different types of regions, including the hierarchical systems, and those bounded by "iso-lines" using "point" data. It suggests objective procedures of delineation of boundaries, that automatically exposes a hierarchy of regions, fixed in number, size and shape, which is already structurally present. It outlines a method of "estimation" of boundaries of hierarchical regions, based on sample survey data, and indicates a visual mapping method of assessing the uncertainties in the estimated boundaries. It sets forth a scheme which is flexible enough to produce, with comparative ease, results on selected indicators, for more than one system of grouping of units into regions serving different purposes. It narrates the attempts at standardisation of regions, for continuous use over rounds, in order to extend and reinforce the inherent advantage, normally restricted to a single round, of the multi-subject design of making available a variety of data for the same set of regions, namely, those used in that particular round. It describes how in spite of its inability to furnish village statistics for each and every village in the country, it can draw valuable regional pictures of the village conditions. It indicates how the survey design has provided for research data to study relationships between different regional peculiarities; and between such attributes on the one hand and village characteristics on the other; and, in turn, between the latter and household or individual characteristics on the other.

INTRODUCTION

1. *Preliminary observations.* The Indian National Sample Survey (NSS) has been carrying on enquiries on various socio-economic subjects on a continuing basis since 1950. Enquiries on consumer expenditure, employment and unemployment, population, housing, household enterprise, consumer price, village particulars, crop acreage and yield etc. have been taken up more or less on a regular basis. Subjects like land holdings, conditions of agricultural labourers etc., have also been taken up from time to time. Production of estimates at the national level has naturally received the greatest attention, but the case for statistics for regions, that is, for areas smaller than the whole country, has not been lost sight of. In fact, in the very first round (1950-51) of the survey, estimates pertaining to six zones constituting the country were obtained. And at the present time, production of estimates for the constituent States is considered an essential objective of the NSS. Even for regions smaller than the States, estimates on some selected items, have been actually obtained in some of the rounds.

* This paper originally published in *Bull. Int. Stat. Inst.* XLI, is being reprinted here with the kind permission of the International Statistical Institute. —*Editors.*

2. This long occupation with the question of regional estimates has naturally encouraged some thinking, followed by some action programme. The whole process is still under development even though we think, we have made substantial advances. It is therefore believed, that it might be of interest to recount here our thoughts on this subject, and also to give an account of the action so far taken.

3. *Regional disposition and role of NSS.* Regional statistics may be considered to fall into different categories, depending upon the purpose; and the NSS has to take account of its general survey conditions in order to decide its role in relation to each type. We have first the category where data relating to only one or a limited number of specified regions are desired for some peculiar problems which are not of interest, at least not urgently so, to the remaining regions. The NSS organisation has on a few occasions produced such regional statistics, either, because of the absence of a more suitable agency in the central region, as in the case of survey of two-roomed houses in Delhi, 1960; or because of wide-spread scatter throughout the country of the regions or centres of interest, and where coordinated plan of work is desirable, as in the case of Family Living Surveys (1958-59) of middle class (45 urban centres), and of the working class (50 industrial, mining and plantation areas). But, being essentially an organisation of country-wide coverage, production of statistics restricted to selected regions only is not generally favoured. The broad category of situation, which is more in keeping with the NSS set-up, is the one, in which in addition to national estimates, results are also required separately for each of the regions into which the country has been divided. National estimates together with estimates for some specified regions, like the Calcutta Metropolitan or Damodar Valley Project areas, are also conceivable. This applies to the situation where regional statistics relate to identical subject fields, which must be of all-India importance also.

4. *Enumerative and research objectives.* For regional statistics, as for others, there are two classes of survey objectives. In one, which may be termed enumerative, summary information on several items is obtained, for direct use in policy formulation and plan implementation. In the second, which is of an investigational or research nature, the emphasis is on making comparisons and on studying relationships for ultimate use in better planning and correct policy formulation. The NSS is in a position to throw up useful data for both the above objectives.

5. For policy formulation, planning needs or action programmes, summary results by the general administrative divisions of the country, like States and districts, are a common requirement. Now, there may be two points of view, differing only in emphasis. To an authority directly concerned with a particular territorial division the characteristics of that region are of definite importance, but comparison of this division with other divisions, may not be of so great significance; whereas, the opposite may be true for one concerned with laying down relative priorities among divisions or identifying the problem divisions. Again, results for zones not co-extensive with the administrative divisions may be desired; for example, in a case where the units for

planning, for some specialised subject, are different. Sometimes the appropriate regional division may not be known prior to survey designing, especially where the data are to be put to use for an additional purpose not primarily visualised.

6. Investigational studies on relationship of different types are of particular relevance in regional analysis. One tries to establish (a) relationship between different regional characteristics, — here the region is the unit of analysis, (b) relationship between the characteristics of non-regional units (for example, housing characteristics), with the characteristics of the region in which such units are located; or that between regional and sub-regional (say, village) characteristics; and (c) the manner in which regional characteristics affect the nature and strength of relationship between various sub-regional or non-regional characteristics. Moreover, the choice and delineation of regions itself may be a subject of investigation.

7. The primary role of the NSS is data collection, and it has been continually on the look out for more and more effective survey designing, so that, within the budgetary limitations, it can produce and provide for as much useful regional statistics as possible, and can at the same time, plan in a manner that a large measure of investigational types of studies may be possible. Our ideas and actions are described below; it is convenient to deal with large, medium and small-sized regions separately.

Large regions

8. *First set of regions—population zones.* Principles governing the choice of regions for the first few rounds of NSS were, (a) precedence and general acceptability, (b) availability of background information for interpreting the survey results, and (c) adequacy of sample size for obtaining sufficiently precise estimates. For purposes of analysis of census data, the 1951 Population Census authorities defined six Population Zones, which covered the entire country. These were formed by grouping contiguous States, lying notionally in different directions. NSS adopted these Zones, and Zonal results on consumption pattern, household enterprises etc. obtained by NSS could then be viewed against the basic population and other statistics, compiled for these Zones by the Census authorities. Some other statistics, although not already compiled by these Zones, were obtainable, because State-wise data were usually available. This is one of the points in favour of forming regions by grouping units, for which official statistics are available. The NSS sample size for a Zone was considered fairly adequate. In the second round, for example, the number of sample villages in a Zone ranged from about 100 to 300, with the number of sample households ten times as much.

9. *Multiple regioning system.* While, in a sample survey, sample size is a serious limitation on the number of separate regions for which sufficiently precise estimates are obtainable, one must remember, that this is particularly so, only when a single system of regioning is considered. In fact one may obtain estimates for

numerically more, but overlapping, regions when, by using different principles, two or more systems of regioning are simultaneously visualised. As an example, we may cite the adequacy of even the smaller samples of the initial rounds of the NSS, for throwing up estimates, not only for the six Population Zones, but also for the five Natural Regions, also established by the (1951) Census authorities.

10. It is operationally convenient, and to some extent technically desirable, to ensure that strata do not cut across the regions. When stratification variables which are not regioning variables, are available, it is desirable to proceed to stratify in two stages. In the first-stage, basic strata are obtained as the "intersections" of two (or more) systems of regioning, and then, the (non-regioning) stratification variables are utilised for sub-stratification. Elimination of stages may lead to difficulties.

11. Even when tabulations by only one of two possible systems of regions are contemplated, it may still be worthwhile to adopt the scheme of double regioning with sub-stratification, if the secondary regioning variables also happen to be good stratification variables in respect of the subjects of enquiry. This is a method of keeping a convenient provision for tabulation for an additional regional system useful for research purposes. This idea was followed up in our design for the fourth round (1952), as explained below.

12. In addition to the six Population Zones, which were groups of States the Census authorities also delineated five Natural Regions and fifteen Natural Sub-regions, on the basis of physical features, soil type, climate, mineral resources etc. These Regions and Sub-regions cut across State boundaries. It was felt, while designing the fourth round of the NSS, that strata should be formed as to cut neither across these Natural Sub-regions nor across the State boundaries. There were two main reasons for this. First, it was felt, that in addition to the regions, —Population Zones, —for which tabulations on a continuing basis had been already contemplated, there should be some provision for convenient tabulations by Natural Regions, (each of which had a fair sized sample), for research purposes. Secondly, the intention of using the Natural Sub-regions, as one of the intersecting systems of regions, (the other system being the division into States), while obtaining the basic strata, (which are essentially the units called Natural Divisions by the Census authorities), was to obtain more precise estimates for regions of both types —Population Zones and Natural Regions. Population density, which was a stratifying variable, was used subsequently to sub-stratify the basic strata.

13. *Administrative vs. natural regions.* Our experience shows that there is a tendency to favour administrative-unit-based regions to regions which are not so, but are basically of a research oriented type. Thus, even though the Natural Regions were expected, on general grounds, to bring out very clearly differences in consumption patterns and household enterprises, it was not considered important enough to be given preference over Population Zones, for actual tabulations even though

such zones were mere groups of States with really no separate administrative or operational unity. One of the major reasons for sticking to a system of regions is continuity; data in earlier rounds were presented by Population Zones, and therefore for the sake of comparison over rounds these zones were favoured. But continuity of the series could not have been the sole reason for this decision, as subsequent history would show.

14. *Modified set of population zones.* The areas of the 1951 Census Population Zones, although defined in terms of entire States at the time of their formulation, ceased to be so later on due to major reorganisation of States in India. The boundaries of the Population Zones passed right across some of the reorganised States. The NSS was faced with the choice between the planning and presentation of data according to the initial delineation, or, to modify the zones so as to fit in with States reorganisation. The NSS opted for the second system of presentation of data. For survey operations, which were to be carried out in close cooperation with the general administrative set up of State Governments, there were distinct advantages in making the field zones, (for survey operations), fit in with the general administrative set-up; and, making the stratification and tabulation correspond to the field zoning was obviously convenient. The design did provide for not inconvenient tabulations for adjustment to the other system, but such tabulations were not actually made, and strict comparability of regional estimates over rounds was lost. The second system of zoning was considered, however, to be more in keeping with "reality", and therefore more useful.

15. *Second set of regions—council zones.* The Population Zones were merely a convenient grouping of States for presentation of data; and had no status as units for policy formulation or action programme. Another system of grouping, which had some such significance, came into existence later on. These were the zones, five in number, corresponding to certain Councils, set up by the Centre, where representatives of different States, constituting any zone, occasionally met for consultations on questions of zonal nature. To enhance the usefulness of regional estimates, the NSS switched over to these five zones in place of the older Population Zones.

16. *Third set of regions—states.* The NSS sample grew larger and larger in later rounds, and production of usable estimates for more than five or six regions, became a practical proposition. Tabulation of results for a larger number of regions, (States and some Union Territories), became the practice. It must be pointed out that treatment of States as regions was not just an incidental consequence due to expansion of the sample size, but, that expansion itself was mainly motivated towards producing State-wise estimates, for which there were overwhelming demands with the increasing tempo of developmental activities. The States as regions are much more useful than either of the previous types of zones, because the State, which has substantial amount of autonomy, is a very important unit for policy decision and plan formulation and implementation, whereas even the Council zone is merely a grouping for occasional consultations.

17. *Steps leading to increase in sample size.* The major hurdle in the production of regional statistics in a sampling enquiry is the limitation of sample size, and it seems desirable to give an account of the steps taken to meet this obstacle. The major step, completed last, was the participation of the States on a matching basis; we shall speak about it later on (paras 27, 28). There were two other principal steps which were helpful, but, in what sense they contributed to the sample size, needs some discussion. These steps were (a) lengthening of survey period of a round, and (b) increasing the subject or item coverage of a round. To appreciate the significance of these steps, it is necessary to remember that the NSS is essentially a continuing multi-subject survey system, with more or less permanent whole-time field staff; a good part of the data being collected, is obtained by interviewing householders. In any particular round enquiries on a specified set of subjects are carried out, usually in different sets of households in the same visit to a sample village/block by the roving investigator. Some of the subjects are being kept common in all the rounds.

18. *Survey period and sample size.* That a larger sample could be covered by a fixed number of investigators (or enumerators) working for a longer period, is self-evident. But, there were some technical difficulties which had to be overcome before full exploitation of this principle could be made. Our initial way of looking at things had to be carefully reviewed. In the first few rounds, not much attention could be paid to possible effects of different timiings of visits within the survey period. But as a precautionary measure, the length of the survey period or of the round was made comparatively small, so that the effects of such variations in timings might not be large. This restriction imposed a limit to the sample size manageable by a fixed field strength.

19. With the passage of time, greater and greater pressure was being put on the NSS to provide for more and more precise estimates; also, data for larger number of regions, the States, were demanded. The NSS yielded more and more to such pressures by lengthening the survey period, beyond what was considered safe earlier, to fit in with sample size considered necessary. This undesirable feature of the survey plan was brought to the fore-front, during the first serious attempt in the eighth round to provide State-wise estimates, viz., those for land holding, to which a further reference would be made in para 27. The survey period (round) had to be made longer than that for any previous round in order to accommodate a larger sample. Although all the States did not come for regular participation on a matching basis in subsequent rounds, one thing was clear that there was no going back from the idea of State estimates, and in the following rounds prolongation of survey period for larger samples became necessary. A method, therefore, had to be evolved for controlling possible biases due to uncertainties in the timing of visits during a long survey period.

20. *Use of short moving reference period.* There would have been no difficulty with the older survey arrangement if the informants supplied data from records maintained by them. But this was more the exception than the rule, so that, the data

relating to a fixed calendar period or season would be subject to different recall periods, and data relating to a rather old period were likely to be less reliable. This question of recall lapse was sought to be controlled by having a short moving reference period, whose length was fixed but whose end-points were moving. The investigators collected data, for example, for the day, week or month immediately preceding the date of their visit to a household.

21. *Necessity of sub-rounds and space-time representation.* While the moving reference period proved helpful in reducing the effect of recall lapse in a survey plan with a long survey period there was another connected problem with which we had to contend. With the earlier survey plan, the field organisation decided, according to their operational convenience, as to the sequence in which an investigator would cover the sample villages in the area to which he was assigned. The part of the sample covered in any sub-period (of the survey period) was, no doubt, well scattered throughout the country, but one could not strictly claim that this constituted a probability sample for the sub-period, even though the totality of the sample villages, planned to be covered in the full round, constituted a probability sample of villages. Our objective was to secure an unbiased average picture over the entire survey period and therefore, data relating to every sub-period should have to be an unbiased sample for the sub-period. The total sample was unbiased in respect of space representation, but data failed to be strictly unbiased as regards space-time representation. However, with a short survey period, the distortion is likely to be insignificant. With a long survey period, on the other hand, the distortions may not be negligible. It should be noted that the whole-time investigator has to carry out the survey work continuously in one place or another without any break; this possibly makes the bias much less than what might have been if the investigator worked intermittently with freedom in the choice of the period of his visit to a village.

22. The solution already hinted at in the previous paragraph, was to break up a long round into a number of sub-rounds of short duration, and then, chose a probability sample for each sub-round. Once this is done, one can, by increasing the length of the round, accommodate samples of any size. This system of sub-rounds has been in vogue in the NSS for nearly ten years. The length of a sub-round is currently two months; (smaller sub-rounds were also adopted in some of the earlier rounds). Also, unlike the earlier practice, the field-staff are not all free to choose the order of visit to sample villages in a sub-round. The order is fixed according to the principles of probability at the survey design stage. Thus, this device made the sub-rounds, in effect, smaller still.

23. *Subject multiplicity and field strength.* Increase in the multiplicity of subject coverage required additional investigation time, and therefore with a fixed investigation strength it needed lengthening of survey period. Elaboration of any subject of enquiry with the introduction of new items of information had also the same effect. Supply of additional results itself created more demand for data, and for

more precise results. Elaboration of an old subject or introduction of a new subject of enquiry, which is possible in a multi-subject survey organisation, was considered a good ground by the authorities for sanctioning additional field staff later on. And as an investigator has to take up all the subjects of enquiry simultaneously in his visit to a sample village, there was no question of setting apart separate batches of investigators for different subjects. Thus, the increase in the subject or item coverage, helped the NSS to reap the benefit of increased number of investigators for all the enquiries, and also for regional (State) estimates.

24. *Subject multiplicity and sub-rounds.* Multiplicity of subject coverage (and multiplicity of regions) initially augmented the difficulties concerned with lengthening of survey period but came ultimately to its rescue by inducing the authorities to sanction more field staff. Possibly, the system of sub-rounds could not have been introduced without further difficulty, if the NSS had not been a multi-subject survey organisation. This requires some elaboration. Here it should be pointed out that the introduction of sub-rounds introduces greater travelling by the investigators, because they have to cover the entire area in every sub-round. This may not be a feasible proposition, unless the investigation zone, (that is, the area to be covered by an investigator), is sufficiently small. Now, in our current system, the geographical stratum is made the investigation zone of an investigator in the rural areas. (In practice in every investigation zone four investigators are posted, two belong to the Central staff and the other two to the State staff). This means that the size of the investigation zones is inversely proportional to the total number of investigators. With increase in field staff due to increased subject multiplicity there was decrease in the size of investigation zones. This in turn enabled implementation of the scheme of sub-rounds, which was not feasible earlier.

25. *Sub-rounds and seasonal variation.* It may be mentioned, in passing, that the need for lengthening the round was also felt on other grounds. The economy depending as it is on agriculture mainly, is quite seasonal, and to eliminate the seasonal effects, it is desirable to average over a full year. Thus, in later rounds the length of a round was made one year, coinciding exactly with, what is known as the agricultural year, (July-June), in India. At the same time there must be provision for seasonal estimates, and the scheme of sub-rounds met this point.

26. *Subject multiplicity and sampling efficiency.* Increase in multiplicity of subject coverage, made a direct contribution to the increase in sample size. Savings on common operations like journey, arrangements for boarding and lodging, household listing etc., could be utilised for increasing the sample. It could also make an indirect contribution which, in effect, amounted to an increase in sample size. This flows really from certain features of the concurrent multi-subject survey system. To understand the position, one must realise, without going into greater technical details, that integration of different enquiries enabled the survey on different subjects to be spread over a very much larger number of sample villages, than what would have been possible,

with the expenditure of the same resources, if the subjects were taken up separately by different batches of investigators. It is also known that a household sample of given size spread over a larger number of villages, is equivalent to a much larger sample of households spread over a substantially smaller number of villages. Thus the multi-subject system indirectly helped, in effect, to increase the sample size.

27. *State participation.* Perhaps the most effective step was to convince, and then, to encourage the States about the worthwhileness of spending some of their resources, in augmenting the sample size to help production of reasonably precise State estimates. The first opportunity for such a joint effort came when India participated in an international effort, viz., The World Agricultural Census in the 1950's. Being predominantly an agricultural country, and noting the importance attached to it by the Food and Agriculture Organisation of the United Nations, all the State authorities considered it worthwhile to join hands with the Central Government in the survey of agricultural holdings, on a sample basis. In the eighth round (July 1954–March 1955), the NSS Organisation of the Government of India collected data from some sample units and the States collected data on additional sample units, which were, however, tabulated by the Central agency.

28. While the outcome of this action no doubt gave concrete evidence of the value of joint effort, it was some time, before the majority of the States came in for regular participation in the different rounds of the NSS. The Centre encouraged the States in their participation by meeting a part of their survey costs. This encouragement immediately bore fruit, and from the 9th round (1955) two of the biggest States in India started participating in the NSS on a matching basis. The general plan is that a State should canvass in a sample, which is drawn in exactly the same manner as the sample covered by the Central agency in that State. The State follows the same concepts, definitions and methods, and processes their own data independently, but adopts identical estimation procedure. The State sample is usually of the same size as the corresponding Central sample. The State, however, is free to have a sample size which is any convenient multiple of the Central sample size for the State. Such freedom has been exercised by some of the States from time to time. This does not mean that the State has to augment its sample size for all the subjects canvassed; it can do so for even one subject only. There are two aspects of this arrangement. It meets the point that the sample size for a State, which is reasonable for national estimates only, may not be so, for State estimates. And what is more, it is a test whether the permissible error demanded by a State is really essential; that is, whether the State considers it so important as to incur extra expenditure for increasing the sample to a size which is in conformity with the permissible error specified by her.

29. *Changes in sample allocations for state estimates.* It is true that the States are vitally interested in the estimates relating to their respective States. But the Centre is also interested in such estimates; and, therefore, it is only proper that the responsibility for obtaining precise State estimates should not be left entirely to the

States. Moreover, all the States did not join the NSS simultaneously. But once it was realised that in addition to national estimates State estimates are to be produced, we did not wait for a particular State to join, but quickly started alteration of field-strength and therefore of sample size in the different States, so that even if precise estimates are not obtainable, tolerably useful ones would be, at least of some important characteristics. The change was in favour of the smaller States. Scope for adjustment without discharging any one was possible due to expansion in field strength and also to some turnover everywhere. In a survey organisation, where greatest importance is attached to the accumulation and exploitation of the experience gained by individual investigator (or enumerator), nothing can be more undesirable than terminating the services of a competent and experienced investigator in one State, while making recruitment in another State of a new-comer, transfer being usually not possible due to language difficulties. The process of change-over was therefore gradual. At present in a single round no State claims less than 1000 sample centres (villages and urban blocks). The biggest State (Uttar Pradesh) has about 3200 such centres.

30. *State participation and comparability, validity and reliability of results.* The Central Government sought the cooperation of all States, not only for increasing the sample size, but for other weighty reasons also. In India the constitutional set-up is such that, in many subjects, the States are more or less independent in the matter of collection of statistics. One major purpose of the production of State estimates is to ensure the availability of comparable data, and, therefore, the matter of coordination between States becomes rather crucial. It is not merely a question of coming to an agreement about uniform concepts, definitions, data collection procedures, and set of end results, in case the subjects under consideration are taken up by a State; but, one has to ensure that such investigations are actually taken up by the States. One has also to ensure, that the various States take up simultaneously investigations on the same subject, so that the individual results relate to the same time period. The NSS with State participation ensures all the above requirements.

31. Another important consideration, in a sense more important, which has dictated the participation of the States in the NSS, is related to the question of validity and accuracy of results. Keeping the question of accuracy in mind, a survey organisation (NSS) with permanent whole time staff with adequate supervision at every stage, was visualised. Even taking all these precautions, in a large scale survey one needs some evidence that the work has been done under proper control. Such evidences are thrown up by the data themselves. We have previously mentioned that the State authorities and Central authorities obtain and process the data independently, and therefore two completely independent estimates are available, —one provided by the Centre and the other by the State. In fact, the Centre as well as the States each provides (at least) two estimates for the same characteristics, the data, relating to two sub-samples, being obtained for each agency by two different batches of investigators. These estimates

enable us to assess the margin of uncertainty and the quality of data. If there is much divergence between the two estimates thrown up by the different agencies, then there is *a priori* reason to doubt the quality of the results, and an investigation is usually made to detect the causes of this divergence. It has not been an uncommon experience to discover the existence of some processing errors, and sometimes faulty understanding of the instructions, and even faulty instructions themselves. Finally, the participation of the States helped to draw up a more useful programme of work, because the State requirements and priorities have better chance of being appreciated with their participation. Also, their experience and suggestions have been of considerable help in improving the survey plan.

Medium sized regions

32. *Zones—agricultural labour enquiries.* We have already given an account of our efforts leading to State estimates, and we shall now confine our observations to sub-regions of States that is, regions within States. The first all-India attempt at such regional estimates in a sample survey was made in the First Agricultural Labour Enquiry (ALE), 1950-51. It was found advantageous, both from administrative and statistical considerations, to divide each major State into a few homogeneous zones for the purposes of the enquiry. The zoning was carried out on broad agricultural and economic considerations, and on the basis of judgement and knowledge of the State officers concerned, like the Directors of Economics and Statistics, Directors of Agriculture, Directors of Land Records etc. The country was divided into 75 zones. While the zone in a few cases cut across district boundaries, they were so demarcated as not to cut across *tehsil* boundaries.

33. The Second Agricultural Labour Enquiry (1956-57) was conducted by the NSS Organization of the Government of India, and by only four participating States. The country was ultimately divided into 38 zones by suitably combining the 75 1st-ALE zones. This combination was considered desirable so that the sample size may not be too inadequate for a zone. The average population of a second ALE zone was somewhat less than ten million. The average number of sample villages per zone (excluding the sample covered by participating States) was nearly 95 against 11 in the first ALE, and the average numbers of sample households (with a single visit by the investigator) were more than 200 for enquiries on consumer expenditure, and about 550 for employment, against the earlier size of nearly 150 (with monthly visits) for either type of enquiry. The minimum number of sample villages in a zone was 40 against 2 in the earlier enquiry. Being part of a multi-subject survey, it was economically feasible to spread the sample agricultural labour households over a larger number of villages, and this helped to increase the precision.

34. *Regions—landholdings enquiry.* The second occasion, when regional estimates were planned for, was in connection with the World Agricutural Census around 1960. The Government felt that it was an opportunity for obtaining regional estimates. In consultation with the States and concerned Central ministries a system of 48 regions, obtained by grouping of districts, was evolved. Here the starting point

was the set of Natural Divisions defined in the 1951 census (para 12). With some modifications, made mainly on the basis of latest available information on crop pattern and population density and some general knowledge of competent persons in the States, the final regioning scheme was reached. Fullest advantage was taken of the States' participation in the NSS. In a region there were on the average about 300 sample villages and 4000 sample households.

35. *Need for pooling over rounds and years.* The first time when a deliberately planned attempt was made in the NSS to combine the data for two successive rounds, was in the Second Agricultural Labour Enquiry, 1956-57. To increase the sample size, different sets of villages were selected in the two rounds, which were each of six months' duration. Another reason for combining, was to secure a picture averaged over a year, because of seasonal nature of employment in agriculture. Important items on which zonal estimates were provided were annual income per family, consumer expenditure per family, average number of days employed/unemployed, average number of days worked as casual labourers in different agricultural operations in a year, etc.

36. The idea of enlarging the sample by extending the survey period even beyond one year, was actually implemented in connection with regional estimates for the landholdings enquiry, 1960-62. A large sample was needed for regional estimates, but a temporary increase in field strength for this purpose was not desirable. It was decided, therefore, to depend entirely on the regular experienced staff, and spread the sample over the two consecutive rounds, 16th and 17th, each of one year's duration.

37. *Standard regions and multi-subject surveys.* There are obvious interpretational and other advantages in having a variety of data on several related subjects tabulated according to a single system of regions. In a multi-subject survey, where several subjects are taken up simultaneously in a round according to an integrated sample design, the use of a common regional system for all the subjects is almost automatic. And if we could stick to identical common regions, round after round, then the results would be more useful. This we have been striving to achieve for a fairly long time. Considering the fact that the country is basically agricultural, and that the regioning characteristics, used in the 16th and 17th rounds (1960-62), were of a general nature, and also remembering that this NSS region system has received the approval of various State governments, it was felt that this system of regioning might be adhered to in subsequent rounds also. We have been able to hold on largely to this NSS system for the next two rounds.

38. Our ability to stick on to the same regions for three consecutive rounds has proved to be rather fortunate, as explained below. There have been instances, where the need for regional estimates was felt subsequent to the completion of a survey. The latest instance is the one, where the Planning Commission require data to study the regional disparities in connection with the formulation of the Fourth Plan. The

NSS regions, established earlier, proved handy, and are being utilised for this purpose. Use is being made, not only of the 16th and 17th round data on land holdings, which motivated the NSS regioning, but also, of the 18th and the two earlier round data on household consumption and employment. These regional results are expected to throw some light on relative levels of living.

39. Our attempts at standardisation have however been only partially successful. In fact considerable pressure is now being put on the NSS by some of the States to alter their regions. Also, subsequent to NSS region formation, the Resources and Scientic Research Division of the Planning Commission have proposed after very elaborate studies a system of Resource Development Regions and Divisions of India. They recognize, that for certain planning purposes, this system may not be appropriate; but they feel that, for a large area of planning, acceptance of their scheme would be a great advantage. The scheme is based on consideration of physical factors—topography, soil, geologic formations and climate; and agricultural land-use and cropping pattern. There are fourteen Regions demarcated mainly on physical factors; most of them cut across different States, but the Divisions which are fiftyseven in number do not. In forming Divisions account has been taken not only of physical conditions but also of agriculture. Regions and Divisions have been formed by grouping districts. With the availability of further statistical information, and the suggestions of various interested bodies, including the above regioning of the Research Division of the Planning Commission, the question of NSS revising its regions has arisen. Only time will show to what extent the revision (yet to be decided) will last.

40. *Need for objective methods in region formation.* It is very difficult to arrive at enduring unanimity in a multi-agency, multi-subject survey system, in spite of the fact, that there is a general appreciation of the value of sticking to a single standard system of regioning. This is not unexpected; the regioning partly depends upon the purpose for which a particular agency wants the regions at a particular point of time, and partly on the experience and judgement of the individual who represents the agency on a particular occasion. Difficulties arise in matters of details, like number of regions, and their exact boundary specifications. It should also be remembered that regioning not only depends upon the subject of interest, but also on the level at which one is formulating a policy or evolving an action programme. I think, it is important to minimise the subjective element in the formation of regions by laying down well-thought-out objective procedures.

41. *Choice of units for regional grouping.* We shall now concentrate on different aspects of the problem of formation of regions. The first question to be decided is the type of unit (territorial division), which should be grouped together to form a region. Information on the geographic position of units is essential; for, the region, as ordinarily conceived, must be a geographically compact area. But in order to make a rational choice among innumerable possible systems of subdivisions of a State into geographically compact clusters, one must have some useful information, regarding the

units to be grouped into such clusters. Keeping this point in view, one can think of three important types of units in India : (a) district (population more than a million on average), (b) *tehsil* (population about 150,000) or (c) village (population 600 on the average). Recently, the entire country has been covered by Community Development Blocks. As units, these are as yet not quite suitable because of lack of readily available and dependable data for each such block, roughly half of a *tehsil* in size on the average. Insistence on using these blocks as units may grow in future.

42. It is not at present a feasible proposition to think of paying individual attention to the village particulars, (there are more than half a million villages in India), for the formation of medium-sized regions. So, we are left with the choice between the district and the *tehsil*, which are much more fundamental than the village in the administrative set-up. Also, more statistical and other types of information are readily available for these units. There are advocates for each of these types of units. Those in favour of the district argue that it is more important, being higher up in the hierarchy of administrative divisions of the States, and somewhat more information (for purposes of grouping) is readily available; and, therefore, it has a greater claim for consideration as the building unit of the regions. The opponents argue that the district is much more heterogeneous than the *tehsil*. Also, the latter, being more numerous, offers a greater scope for construction of more homogeneous regions.

43. *Types of regions.* Even if the question of the building unit is resolved, we would be still left with a number of difficult questions. We have first the question of choice between the two types of regions, which seem to have gained some recognition—(a) the homogeneous and (b) the functional or nodal. The first takes cognizance of the similarity between (adjoining) units, and the second, the strength of functional relationship between units. Sub-division of the entire country into nodal regions may be practically ruled out; because in India very little statistical information is readily available on the various types of flows or transactions between any two sufficiently small units, except to the minor extent described below.

44. The NSS investigators and supervising staff have had several years' experience of travelling, from a random village in one *tehsil* to another in a different *tehsil*, and are in a position to assess the difficulties involved. Statements, showing their assessment of the different degrees of difficulty or ease of travelling, were obtained for each pair of adjoining *tehsils*. Also, their assessment of the relative ease of movement from a given *tehsil* to all the adjoining *tehsils* was obtained. Strata formation, by grouping contiguous *tehsils*, took the ease of travelling as one of the criteria for grouping, and regions were formed by combining a few strata. The stratification and regioning were functional to the extent the ease of travelling reflected the strength of economic bond.

45. *Concept of "closeness" and principle of regioning.* Whatever be the conceptual difference between the two types, as an exercise in formation of regions, the difference is probably not as much. One has first to devise a suitable measure of the

"closeness" between the two units; in case of homogeneous regions "closeness' measures the degree of similarity between, and in case of the functional or nodal region "closeness" measures the volume of flow between, or the strength of the bond connecting two units.

46. Measurement of the "closeness" or "distance" between two units has two aspects. First, one has to enunciate the principles on which the choice of the characters or transactions for consideration of "closeness" is to be based and secondly, one has to improvise a composite yard-stick, for measurement of "distance", when more than one character or item are selected.

47. Even when the above questions are resolved, one has to evolve a meaningful objective procedure of determining the number of regions and delineating their boundaries. For a prescribed number of regions, one reasonable principle would be to pick up that system of regioning which minimises the sum (simple or weighted) of the "inter-member distances" of all such pairs of units, whose members belong to the same region; that is, choose the one which maximises the sum of the "inter-member distances" of all such pairs of units, whose members belong to different regions. Some simple procedure is to be evolved for arriving quickly at the desired system of regioning.

48. *Hierarchy of regions.* The following gives a process of setting up a hierarchy of regions which is invariant under any order preserving transformation of the chosen measure of "distance or closeness". We shall, for the sake of simplicity, assume that no two pairs of adjoining units have exactly identical "inter-member distances" between them. With any such pairs the only modification necessary in the procedures described below is simultaneous removal of, (or erection of barriers along) the common boundaries of such pairs. First we work from the maximum number of regions, viz., the total number of units, each regarded as a region, to gradually smaller number of regions. First, remove the common boundary between the two "closest" units; we get now a set with one region less. Next apply the same procedure over and over again. By removing the requisite common boundaries between the "closest" regions of one set, we arrive at the next set in which the previous set is embedded. The resulting hierarchy of regions is completely "structural", and after the choice of the yeard-stick for measuring "closeness" nothing is left to judgement. It will be noticed that the "slope" or "change" between two adjoining units belonging to adjoining regions, of any stage in the hierarchy, is more abrupt than the "slope" between either of these units and any of its adjoining units (not separated by an unremoved common boundary) belonging to the same region.

49. One can also start from the opposite end, by erecting barriers across the common boundaries. We suppose that the entire country or State, which is being subdivided, is already walled up. Here, we start from the boundarie between adjoining units which are "furthest" from each other, and then the next "furthest" boundary and so on, until we have a region enclosed fully by barriers. At the moment of completing an enclosure, there may be some odd or isolated barriers not enclosing any area. We

proceed next to form other regions or sub-regions by continuing the same process. Perhaps, the most significant sets of regions are those, which divide the country or State into a small number of sub-divisions. Thus, the most significant sets appear in the first few stages in the above process of regioning. In the previous method they appear last. It is worth noticing that members of these sets are large on the average, but they may have very unequal number of units, and may be very dissimilar in shape. Among the regions corresponding to different stages of the heirarchical system special attention might be deserved by those (particularly the ones, if any, with small number of regions), with not even a single barrier which does not form part of a completed enclosure.

50. It will be noticed that at any stage of the hierarchical system, the number, size and shape of regions are automatically fixed. These regions are structural, and each can be considered as a single organic unit like the catchment area of a river, the water-shed being the boundary. The regions obtainable by the method given in para 49 perhaps approximate most closely to "catchment areas" when greater "distance" stands for smaller volume of "flow" between units. Strictly, it should be that stage of hierarchy, where the "flow" across the border is really marginal. Finally, it would appear desirable to give regional estimates for all the important stages in a hierarchical system.

51. *Heterogeneous and functional regions.* There is some evidence in the literature to show that the boundaries of nodal or functional regions cut right across those of homogeneous ones. It would be interesting to examine the circumstances under which it is useful to have the most heterogeneous system of regions, which naturally cut across the homogeneous ones. Heterogeneous regions can be obtained by the principle enunciated in para 47, with the words 'minimises' and "maximises' interchanged. Such regions are also obtained in the manner described in paras 48 and 49, with the alteration that merging takes place where a barrier was erected, and *vice versa*. (It is supposed that "distance" or "closeness" has reference here to similarity between units and not to strength of economic bond). If dissimilar units are more likely to be functionally related to one another, then heterogeneous regions are likely to approach the functional ones. Whether to use homogeneous or heterogeneous regions depends upon the purpose. It will be noted that the collection of homogeneous regions is more heterogeneous than the collection of heterogeneous regions. (Heterogeneous clusters are usually to be preferred in cluster sampling).

52. *The commoner type of regions.* In the type of region which is commonly used, the region is not itself so much a natural unit, but a collection of smaller units, the regional boundaries having a certain amount of artificiality. And some modification of the boundaries, or even in the number of regions, may not change the essential character of such regioning. These component smaller units are held together by virtue of geographical compactness, and certain amount of homogeneity (or occasionally by strength of economic bond). The administrative divisions of a country

can be considered as an intermediate type of region. An administrative division must certainly be considered as a single unit for purposes of administration, but it has otherwise a certain amount of artificiality in the boundaries. Action units, like the smaller administrative divisions, need not necessarily be suitable as units of planning. Some commoner type of homogeneous regions constituted of districts have been actually proposed as planning units (para 39). The use of "structural" (homogeneous or otherwise) regions as planning units deserves consideration. Whether any particular type of structural set of regions would serve a variety of planning needs is difficult to say without close examination. Perhaps, all types of structural regions have their uses.

53. If homogeneous regions are formed by grouping contiguous action units, then action-takers or policy-makers, in the absence of separate figures for the respective units, sometimes assume as a first approximation that the characteristics for any constituent unit are very similar to those for the homogeneous region as a whole. It is difficult to say to what extent this assumption is justified. In this connection one must remember that for such use the idea behind the formation of homogeneous regions may be, not just to ensure uniformity (or rather to control heterogeneity) in respect of the characters for which prior information is available, but also to secure uniformity to the extent possible in respect of the items which are not taken into account in setting up the regions, and which are (usually) to be estimated by a survey. If the regioning characters were extremely highly correlated with these items then one should try to achieve maximum homogeneity in respect of the regioning characteristics, by giving up the restrictive condition of geographical compactness. In other words the idea of formation of regions, as ordinarily understood, may have to be given up. But this is rather an exception than the rule; the correlation may only be rather mild, if not low. If the characteristics to be estimated, happen to be those, which are such, that contiguous areas are likely to be more similar in such characteristics, then contiguity is an important factor also to be taken into consideration. Therefore it appears desirable that a large area, even if it be homogeneous in respect of the regioning characteristics, should be sub-divided further into smaller compact areas for separate estimates. But what area should be considered too large requires special study. On the other hand, if the correlation over space is low, and the relationship with the regioning characteristics only mild, then there may not be much point; in fact, it may be rather undesirable to have too many regions.

54. *Continuing survey and district estimates.* We have so far been concerned with the construction of regions, which are groups of districts (or *tehsils*). We may now consider the possibility of treating the district itself as a region. There is now a demand for data, separately for each of these administrative divisions, more than three hundred in number. The urge for district statistics has been increased by the participation of States in the NSS. The district was certainly too small a region (population of somewhat more than a million on the average), for which NSS

could hope to produce usable estimates for any of its earlier rounds with its small field strength at that time. With the passage of time, however, there has been gradual expansion of trained and experienced field staff. And we have perhaps reached a stage, where with about 16,000 sample villages and 9,000 sample blocks in a round, it is not impossible to think of district estimates showing the conditions averaged over a few years (rounds). There are, however, two conditions which, fortunately, are not impossible to fulfil in a continuing survey. We have first to ensure the continuing collection of information on some key items. Although no policy decision has yet been taken, there is some possibility of continuation of the NSS "Integrated household schedule". This tries to give an integrated picture of the household activities and other particulars. It covers demographic and housing particulars, household consumption and enterprise particulars, migration and employment details, etc. At the same time, there is some possibility of continuation of collection of statistical and other information, relating to the village/block in which the sample household resides. The second condition is that there should be replacement, at least partial, of sample villages and blocks over rounds, so that the effective sample size, after pooling of rounds, becomes substantial for a district. It is true that district estimates would involve waiting for several years, but this is a necessary evil with which we are accustomed, when one remembers about the district population data provided by the population census once in ten years

55. *Value of roundwise compilation of district and tehsil figures.* I feel that, for more than one important reason explained in the next paragraph, compilation of roundwise district and even *tehsil* figures as a routine matter, should be given serious thought, in spite of the fact that estimates for any single round are likely to have large sampling errors, particularly for a small region like *tehsil*. To make such a scheme of reasonable dimensions, one cannot think of a full-fledged tabulation at such small territorial levels. Very careful selection of a few items, which can be regarded as indicators, is to be made. The (summary) information, viz., aggregates of all households or other units in a district or *tehsil* should be suitably transferred (to punched cards) for tabulations as and when necessary. For *tehsils*, with our present sample size, there may not be much scope for summarisation. But for districts one can think of summarisation, and therefore, one can possibly think of including a large number of indicators.

56. This scheme serves several purposes, one of which has already been indicated earlier, viz., obtaining district estimates by pooling the data of a few consecutive rounds. And once such estimates are available, one can utilise them for setting up good district-groups, to be treated as regions for every subsequent round. There is another very important practical application, for which waiting for several rounds is not at all necessary. It has been pointed out earlier that different parties may be interested in different systems of regions, (district or *tehsil* groups), depending upon the subject and also upon the purpose for which results are required, and NSS being in the main a general purpose survey, the same general indicators might be

considered useful for different purposes which require different regioning systems. Obviously, region estimates (using ratio estimators, if suitable), can be built up from district/*tehsil* indicators, if already complied. Also, these indicators would prove to be an admirable material for research work on the usefulness and relative merits and demerits of different systems of regioning based on non-NSS regioning data.

57. *Value of systematic village/block sampling.* Before passing on to a completely different kind of use of district and *tehsil* figures, it may be stressed that we have distributed the sample villages/urban blocks in a manner, which is expected to reduce the uncertainty in the incidence and in the number of sample villages/blocks falling in a specified district or *tehsil*. To maximise the number of *tehsils* having at least one sample village/block, we have adopted a systematic method of selection within every geographical stratum, in preference to non-systematic methods, without, however, violating the desired probability of selection of a village/block. A stratum is comprised of a number of *tehsils*. These *tehsils* are first arranged in a "serpentine" manner so that contiguous *tehsils* also appear as such in the list of *tehsils*. The villages within a *tehsil*, as given in the frame, are not infrequently arranged in a similar manner. Then villages are drawn with probability proportional to size (roughly proportional to population), but in a systematic manner with a random start. Our current sample of about 16,000 villages is broken up into four sub-samples, and there are some 3,000 sub-divisions like *tehsils*, so that all *tehsils*, with possible exceptions of some very small ones, are represented in every sub-sample. These minor *tehsils* are to be suitably combined with adjoining ones to form modified *tehsils*, so that every *tehsil* (modified if necessary) is represented in each sub-sample. Systematic sampling maximises the number of such *tehsils* and thereby enhances the flexibility in the formation of regions. For selection of urban blocks, the towns within a geographical stratum are arranged by means of livelihood pattern, and then systematic samples with random start are taken.

58. *"Estimation" of region boundaries.* Now we come to a completely different approach to the problem of regioning. Here we do not regard the boundaries of regions as given for use in survey tabulation, but consider determination of such boundaries, on the basis of survey data themselves, as one of the objectives of a round. If a method of sample delineation can be devised, then one can obtain region maps for any item (or groups of items), data on which are being collected. One possible method is as follows: We decide on the unit (district/*tehsil*) to be grouped and also decide on the appropriate measure of "closeness". We then choose a suitable procedure of forming regions in a purely objective way from among the methods outlined earlier. Accepting the sample estimates for a unit (district/*tehsil*), one applies the above procedure and arrives at the region boundaries. To get a visual idea of the margin of uncertainty in the delineation of boundaries, one repeats the same procedure for each one of the interpenetrating sub-samples into which the NSS data are being collected and tabulated. The "distortions" of these sub-sample boundaries from the total-sample boundaries, provides a visual means of assessing the uncertainty. The value of this

method is yet to be fully assessed. Incidentally, with our present sample size, use of *tehsil* data may not be suitable for the above purpose.

59. *Iso-lines as region boundaries.* Another, but similar, method has received some experimental attention. Price quotations were obtained from some 700 "points" scattered all over the country every week, and iso-lines in the manner of constructing them for meteorological observations were obtained for the sub-samples in addition to those for the total sample. Difficulties in the interpretation of the maps arose because of the unconventional nature of the whole approach. Studies and training in reading such maps are necessary. It will be seen from the next section on small regional centres, that NSS is collecting a variety of data at several "points,"—villages or towns,—and some of them may be amenable to iso-line treatment.

SMALL REGIONS

60. *Village/town statistics.* We shall next explain the role of the NSS in respect of regional statistics of very small areas, namely, villages and towns. In India there are more than half a million villages and nearly three thousand towns. The most outstanding statistical publications on such localities are the district census handbooks showing, among others, the population and number of workers with different means of livelihood for each village and town. There is some unpublished information, e.g. at the instance of the NSS the Census authorities compiled village-wise data on number of small industrial establishments.

61. Obviously, it is not possible for the NSS to produce estimates for each and every unit like village or town or market. But with an extensive permanent field organisation it is particularly suited for periodic collection of data from a fixed but limited set of points randomly scattered all over the country. In fact, monthly series of price data are being produced in every round by the NSS for the same set of more than four hundred rural marketing centres scattered throughout the country. Being combined with other activities of the roving investigators, the cost of price collection is only marginal.

62. We have already pointed out that the NSS is at present covering a very large number of villages (about 16,000) in every round. Therefore, in principle, there is no difficulty in collecting and publishing data for each of these villages, and in the course of several rounds the number of villages covered would be considerable. As for urban centres, all the large and medium-sized towns and a substantial fraction of very small ones are visited in each round, and therefore in principle, it is also possible to produce suitable annual data for each and every town with some exceptions from among the smaller ones. In practice, as far as the urban sector is concerned, data are at present not being collected for the town as a whole, but only for the urban "blocks" which fall in the sample. For the sample villages, a large variety of data is being collected. The NSS has, at present, no plans of publishing these individual village statistics.

63. *Summary picture of village conditions.* Now, although the NSS cannot hope to produce information for each and every village in the country, it can nonetheless present summary pictures of the village situation for States or other large regions; e.g., NSS can estimate the proportion of villages with a school, or the percentage distribution of villages by distance (ranges) from nearest medical practitioners. Such summaries are very important for better comprehension of the village conditions in any particular area, and by studying these features over rounds, one can assess the effects of the developmental activities.

64. *Village/block data at marginal cost.* It is always possible to obtain summary village statistics of the above type on a sampling basis, provided the sample villages are sufficiently large in number, and there is an adequate degree of scattering of these samples over the State or other appropriate region. The main difficulty therefore is the limitation about the resources. However, because of its peculiar nature, the NSS can produce village (or block) statistics at a marginal cost. The main reason, why the NSS is in a position to produce sufficiently precise summary village data at a marginal cost, is because of its multi-subject, multi-purpose, multi-agency and continuing nature. The NSS is collecting on a continuing basis detailed information on a variety of subjects. In connection with these detailed enquiries, the whole-time permanent roving investigators have to visit a large number of villages and towns, and have to incur substantial expenditure on journey. In addition, they have to spend a sizeable portion of their time in the listing of households in the sample village or block, for the preparation of the sampling frame to be used for sub-sampling of households for these enquiries. How these activities help collection of village statistics at marginal cost is explained below.

65. *Two classes of village data.* There are two classes of village/block data which are being collected. The first can give the distribution of different types of village communities, as judged from the characteristics of the resident households. The basic data are being collected at the household listing stage. The data collected during this stage varies to some extent from round to round. These household characteristics are household size, number of couples, period of stay in the village, household enterprise, major and minor source of income, size-class of land holdings, wage earning household or not, number of unemployed persons, number of handicapped persons, etc. For the above type of information the expenditure on data collection may be considered to be nil, as far as production of village statistics is concerned. It is true that the investigators have to spend sometime on the collection of various items of information listed above, but such information is required basically for purposes of stratification or phase sampling of households for the detailed enquiries on consumer expenditure, household enterprise, employment etc. In other words, even if there were no programme for obtaining summary information relating to village statistics, these data would have been collected. Incidentally, collection of these ancillary information involves an addition of only about two percent of the total investigation time over the time required for simple listing.

66. We now come to the second class of village statistics, which are in a sense more important. These are not connected with the household particulars, but relate directly to village characteristics. Examples of the type of data which are being collected are as follows : distance of village from the nearest metalled road or bus stop or market or town, doctor or midwife, school and public library; number of students and teachers in various types of schools located within the village; source of drinking water; availability of irrigation facilities; wages for various types of agricultural work; existence or otherwise of non-household economic activities like cooperatives, joint stock company etc. The time spent on the collection of this type of village statistics amounts to only about four percent of the total investigation time taken for all the subjects in the multi-subject NSS system.

67. One of the advantages of having village/block statistics of the above type is that research studies, on the relationship of the household or individual characteristics (demographic, economic or social) with the village characteristics, is possible as detailed data on consumption, enterprise, employment etc., are being collected from sample households in the villages/blocks. One can also relate the village characteristics with regional, —*tehsil*, district or State, —characteristics.

APPLICATIONS OF BESSEL FUNCTION DISTRIBUTIONS

By FRANK McNOLTY

Lockheed Missiles and Space Company, Sunnyvale, California

SUMMARY. The paper presents a concise discussion of some probability density functions which are useful in the treatment of various Military Operations Research problems related to radar discrimination. The presence of some type of radar processor is implied; however, it is of a very simple type whose essential components include a narrow-band filter, a full-wave square-law detector and a post-detection integrator. A very general set of probability density functions are derived which describe the signal and/or noise which is processed by the radar receiver. The functions are useful for both the analysis and simulation of radar cross-section, amplitude, components and phase. As a matter of interest, some of these functions are specialized to the more familiar expressions which one may find in the EE literature.

In an elegant paper Laha (1954) derived many useful properties of a very general distribution function,

$$F(w)dw = (2/\beta)^{P-1}w^{(P-1)/2}\alpha^{P}e^{-\beta/(4\alpha)}\ e^{-\alpha w}I_{P-1}(\beta\sqrt{w})dw \qquad \ldots\ (1)$$

where w, α, $P > 0$, $\beta \geqslant 0$ and P is not necessarily an integer. Here $I_{P-1}(X)$ is the modified Bessel function of the first kind of order $P-1$.

The generality of Laha's function is evidenced by the fact that it includes, as special cases, such distributions as the gamma distribution, non-central chi-square distribution, various confluent hypergeometric distributions and, of course, such simpler functions as the exponential, chi-square and Erlang distributions.

Strangely, expression (1) and its earlier counterparts seem to have been overlooked in the EE literature; although it does include as special cases the probability density functions for :

(1) The sum of n independent Gaussian (mean zero) noise pulses intergrated from the output of a square-law detector. Let $\beta = 0$, $P = n$ and $\alpha = 1/\sigma^2$ where σ^2 is the effective receiver noise power measured in the IF filter (i.e. the variance of the Gaussian noise-voltage distribution).

(2) The output of a square-law detector when the receiver input is given by $s(t)+n(t)$ where $n(t)$ is a narrow-band Gaussian (mean zero) noise voltage with variance σ^2 and $s(t)$ is a signal voltage with constant amplitude A, i.e.

$$s(t) = A\cos[\omega_c t + \theta(t)].$$

Let $$P = 1,\ \alpha = \frac{1}{\sigma^2} \text{ and } \beta = \frac{\sqrt{2}A}{\sigma^2}.$$

(3) The sum of n independent signal-plus-noise pulses (as described in the preceding paragraph) integrated from the output of a square-law detector. Let $P = n$, $\alpha = \frac{1}{\sigma^2}$ and $\beta = \frac{\sqrt{2n}A}{\sigma^2}$.

One can also specialize expression (1) to include some useful situations in which the amplitude A is constant (complete correlation) over n integrated pulses, but fluctuates independently from one set of n integrated pulses to the next set according to some prescribed density function (see (25), for instance). Similarly, special cases of (1) apply to a situation in which the amplitude is constant over n integrated pulses and the radar receiver sums K independent groups of n pulses each (again from the square-law detector output). These latter two applications of expression (1) arise from its relationship to some of the special forms of a confluent hypergeometric probability distribution.

Having established the relevance of Laha's function to the problem at hand, we now write the input to our narrow-band filter, full-wave square-law detector system as

$$s(t) = u(t)+v(t)$$

$$= A(t)\cos[\omega_c t+\varphi(t)] \qquad \ldots (2)$$

where $u(t)$ may represent the vectorial sum of several voltage returns (not including receiver noise) and $v(t)$ is any narrow-band modulation of the carrier frequency ω_c. The starting point of our analysis begins with the assertion that the probability density function (when A is constant) describing a single pulse at the output of the square-law detector is given by

$$F(w)dw = \left(\frac{2}{\theta A}\right)^{P-1} w^{(P-1)/2}\alpha^P e^{-(\theta^2/4\alpha)A^2} e^{-\alpha w} I_{P-1}(\theta A\sqrt{w})dw \qquad \ldots (3)$$

where θ will be taken as unity in order to expedite the discussion which follows.

In order to introduce the concept of random radar cross-section, a general distribution is now assigned to the amplitude A which was taken as constant in expression (3). The distribution for the amplitude has the form,

$$g(A)dA = \frac{(\lambda A)^Q}{r^{Q-1}} e^{-(r^2/2\lambda)} e^{-(\lambda A^2/2)} I_{Q-1}(rA)dA. \qquad \ldots (4)$$

If expression (1) is denoted by $F(w;\alpha,\beta,P)dw$ then (4) would be identified as

$$F((A^2/2);\lambda,\sqrt{2}r,Q)AdA.$$

The choice of (4) was not arbitrary, since it includes (as special cases) several distributions which agree closely with actual data such as the Rayleigh, Maxwell-Boltzman and one-sided Gaussian. Also, when $Q=\lambda'$, $\lambda=(2\lambda'/\bar{A}^2)$ and $r=0$ expression (4) reduces to an amplitude distribution (see expression (25)) which leads to a gamma distributed radar cross-section where the latter includes two of Swerling's distributions (see expressions (28) and (31)).

When the product of expressions (3) and (4) is integrated over all possible values of A (i.e. from zero to infinity) the random output of the square-law detector (for a single pulse) is described by the following probability density function,

$$F_1(w)dw = e^{-\alpha w}e^{-(r^2/2\lambda)}(2\lambda)^Q \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \frac{\alpha^{P+Q+i+j}r^{2j}w^{P+j-1}}{i!\,j!\,(1+2\lambda\alpha)^{Q+i+j}} \cdot \frac{\Gamma(Q+i+j)}{\Gamma(Q+i)\ \Gamma(P+j)}\,dw. \quad \ldots \quad (5)$$

The expected value of (5) can be obtained by straightforward integration and is given by

$$E(w) = (2\lambda\alpha)^Q e^{-(r^2/2\lambda)} \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \frac{\alpha^{i-1}r^{2j}(P+j)\ \Gamma(Q+i+j)}{i!\,j!\,(1+2\lambda\alpha)^{Q+i+j}\Gamma(Q+i)}$$

$$= \frac{(2\lambda\alpha)^Q e^{-(r^2/2\lambda)}}{\alpha(1+2\lambda\alpha)^Q} \sum_{i=0}^{\infty} \frac{\alpha^i r^{2i}}{i!\ \Gamma(Q+i)(1+2\lambda\alpha)^i}$$

$$\left[P\sum_{j=0}^{\infty}\frac{\Gamma(Q+i+j)}{(1+2\lambda\alpha)^j j!} + \sum_{j=0}^{\infty}\frac{\Gamma(Q+i+j)}{(j-1)!(1+2\lambda\alpha)^j}\right]$$

$$= \frac{(2\lambda\alpha)^Q e^{-(r^2/2\lambda)}}{\alpha(1+2\lambda\alpha)^Q} \sum_{i=0}^{\infty} \frac{\alpha^i r^{2i}}{i!\ \Gamma(Q+i)(1+2\lambda\alpha)^i}$$

$$\cdot\left[P\frac{(1+2\lambda\alpha)^{Q+i}\Gamma(Q+i)}{(2\lambda\alpha)^{Q+i}} + \frac{(Q+i)\Gamma(Q+i)(1+2\lambda\alpha)^{Q+i}}{(2\lambda\alpha)^{Q+i+1}}\right]$$

$$= \frac{P}{\alpha} + \frac{Q}{2\lambda\alpha^2} + \frac{r^2}{4\lambda^2\alpha^2}. \quad \ldots \quad (6)$$

Expression (6) may be derived differently by noting that the mean of (3), with $\theta = 1$, is $(P/\alpha)+(A^2/4\alpha^2)$ and then writing

$$E(w) = \int_0^{\infty} wF_1(w)dw = \int_0^{\infty}\int_0^{\infty} wF(w;\alpha, A, P)F(A^2/2;\lambda, \sqrt{2}r, Q)AdAdw$$

$$= \int_0^{\infty} AF(A^2/2;\lambda, \sqrt{2}r, Q)\left\{\int_0^{\infty} wF(w;\alpha, A, P)dw\right\}dA$$

$$= \frac{P}{\alpha} + \frac{1}{2\alpha^2}\int_0^{\infty} xF(x;\lambda, \sqrt{2}r, Q)dx = \frac{P}{\alpha} + \frac{Q}{2\lambda\alpha^2} + \frac{r^2}{4\lambda^2\alpha^2}. \quad \ldots \quad (7)$$

The variance of (5) may be similarly obtained by using the fact that the variance of (3), with $\theta = 1$, is given by $(P/\alpha)+(A^2/2\alpha^2)$.

In order to obtain the distribution of the randomly fluctuating radar cross-section C_x of all reflectors which produce the voltages $u(t)$ and $v(t)$, we assume that the receiver noise is negligible and note that the variable w appearing in (5) is in units of

power and is related to C_x by $w = KC_x$ where K includes all terms in the radar power equation except, of course, the instantaneous cross-section C_x. It follows that

$$E(w) = \bar{w} = K\bar{C}_x = \frac{P}{\alpha} + \frac{Q}{2\lambda\alpha^2} + \frac{r^2}{4\lambda^2\alpha^2}$$

and thus

$$K = \frac{1}{\alpha\bar{C}_x}\left(P + \frac{Q}{2\lambda\alpha} + \frac{r^2}{4\lambda^2\alpha}\right) = \frac{K_1}{\alpha\bar{C}_x}. \qquad \ldots (8)$$

Substituting $w = KC_x$, $dw = KdC_x$ into (5) yields

$$F_x(C_x)dC_x = (2\lambda\alpha)^Q e^{-(K_1/\bar{C}_x)C_x} e^{-(r^2/2\lambda)} \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} (\alpha r^2)^i \cdot \left(\frac{K_1}{\bar{C}_x}\right)^{P+j} C_x^{P+j-1} \frac{\Gamma(Q+i+j)dC_x}{i!\,j!\,\Gamma(Q+i)\,\Gamma(P+j)(1+2\lambda\alpha)^{Q+i+j}} \qquad \ldots (9)$$

where expression (9) is the probability density function for the randomly fluctuating composite radar cross-section of all reflectors and it does not explicitly involve any parameters (other than $\bar{C}_x$) from the radar power equation. Regarding (5) and (9), it is seen that no loss of generality was suffered by setting $\theta = 1$ in expression (3). That is, one may take (4) to be a p.d.f. for θA rather than just A; whence the quantity θ may be easily introduced via the parameters λ and r without any change in the form of (4).

Now returning to expression (2), we write

$$\begin{aligned} u(t) &= A(t)\cos[\omega_c t + \phi(t)] \\ &= A(t)\cos\phi(t)\cos\omega_c t - A(t)\sin\phi(t)\sin\omega_c t \\ &= x(t)\cos\omega_c t - y(t)\sin\omega_c t \end{aligned} \qquad \ldots (10)$$

where

$$A^2(t) = x^2(t) + y^2(t) \qquad \ldots (11)$$

and $x(t)$, $y(t)$ will be referred to as the components of the process $u(t)$.

From expression (4) it can be shown that the characteristic function for either of the random variables $x^2(t)$ or $y^2(t)$ is,

$$Q(t) = \left(1 - \frac{2it}{\lambda}\right)^{-(Q/2)} e^{(itr^2)/[2\lambda^2(1-(2it/\lambda))]} \qquad \ldots (12)$$

whence the probability density function for either $x^2(t)$ or $y^2(t)$ is given by,

$$\begin{aligned} F_s(s)ds &= \frac{ds}{2\pi}\int_{-\infty}^{\infty} e^{-ist}\left(1 - \frac{2it}{\lambda}\right)^{-(Q/2)} e^{(itr^2)/[2\lambda^2(1-(2it/\lambda))]}\,dt \\ &= 2^{(Q/4)-(3/2)}\left[s^{(Q/4)-(1/2)}\Big/\Gamma(Q/2)\right]\lambda^{Q/2} e^{-(r^2/4\lambda)} e^{-(\lambda/2)s} I_{(Q/2)-1}(r\sqrt{2s})ds, \\ &\qquad 0 < s < \infty. \end{aligned} \qquad \ldots (13)$$